Leave-k-out Diagnostics in State Space Models

Tommaso Proietti*

Dipartimento di Scienze Statistiche

Università di Udine

July 2000

Abstract

The paper derives an algorithm for computing leave-k-out diagnostics for the detection of patches of outliers for stationary and non-stationary state space models with regression effects. The algorithm is based on a reverse run of the Kalman filter on the smoothing errors and is both efficient and easy to implement. An illustration concerning the US index of industrial production for Textiles proves the effectiveness of multiple deletion diagnostics in unmasking clusters of outlying observations.

Keywords: Kalman filter and smoother; influence; outliers; structural time series models.

* *Address for correspondence*: Via Treppo 18, 33100 Udine (Italy). Tel. +39 0432 249581. Fax +39 0432 249595. E-mail: proietti@dss.uniud.it. This paper was completed during my visit at the Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin. I would like to thank Jörg Breitung and Helmut Lütkepohl for their kind hospitality. Financial support by SFB 373 is gratefully acknowledged.

1. Introduction

Diagnostics arising from the deletion of one or more observations are a well established tool in regression analysis for detecting outliers and influential observations (see for instance Cook and Weisberg, 1982, and Atkinson, 1985). The case for multiple deletion is that diagnostics built leaving out one observation at a time are unable to detect groups of outlying points. In a time series setting the case for implementing them is even stronger due to the natural ordering of the observations.

Leave-k-out diagnostics for ARIMA models have been proposed by Bruce and Martin (1989). In this paper it is assumed that the data generating process is a possibly nonstationary linear state space form. In this framework, De Jong (1988) and Kohn and Ansley (1989) showed that single deletion or cross-validatory residuals can be obtained by a run of a smoothing filter, supplementing the usual Kalman filter (KF). The paper is concerned with efficient calculation of leave-k-out diagnostics: this requires the inversion of the covariance matrix of the smoothing errors which may be rather large; perhaps more importantly, the off-diagonal elements are not delivered by the smoothing filter and need to be computed off-line.

The algorithm proposed in the paper consists of a set of backwards recursions run on the smoothing errors which parallel the KF computations and diagonalise their covariance matrix, yielding a set of uncorrelated transformed smoothing errors. When initial diffuse and regression effects are presents the filter is augmented by additional recursions paralleling the augmented KF (De Jong, 1991); further recursions are needed to keep track of the changes in the estimate of the initial and regression effects induced by the sequential deletion of observations. The output of the filter thus allows to compute measures of multiple influence on the inferences about initial and regression effects. A remarkable feature of the algorithm is that once leave-k-out diagnostics are computed, leave-r-out diagnostics for $r < k$ are immediately available, so it has to be run only for the maximum k desired. Throughout the paper it is assumed that the hyperparameters are known; in practice, they will have to be estimated before the diagnostics can be computed; thus, the considerations reported in Haslett and Hayes (1998, sec. 4.2) apply.

The paper is structured as follows. Section 2 reviews basic algorithms for stationary state space models, such as the Kalman filter and smoothing filter, and discusses the various types of residuals available for diagnostic checking. Leave-k-out diagnostics are dealt with in section 3 in which the main computational algorithm is introduced. We then turn to state space models with diffuse initial conditions and regression effects (section 4 illustrates the augmented Kalman filter and smoother) and in section 5 we present the necessary extensions of the algorithm to the framework considered. Section 6 provides an illustration with respect to the US index of industrial production for Textiles, and section 7 concludes.

2. Residual based Diagnostics for Standard State Space Models

Let $\boldsymbol{y}_t$ denote a vector time series with N elements; the state space model is

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{Z}_t\boldsymbol{\alpha}_t + \boldsymbol{G}_t\boldsymbol{\varepsilon}_t, \qquad t = 1, 2, \ldots, T, \\ \boldsymbol{\alpha}_{t+1} &= \boldsymbol{T}_t\boldsymbol{\alpha}_t + \boldsymbol{H}_t\boldsymbol{\varepsilon}_t, \end{aligned} \tag{1}$$

with $\boldsymbol{\alpha}_1 \sim \mathrm{N}(\boldsymbol{a}_1, \sigma^2\boldsymbol{P}_1)$ and $\boldsymbol{\varepsilon}_t \sim \mathrm{NID}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$. The system matrices, $\boldsymbol{Z}_t$, $\boldsymbol{G}_t$, $\boldsymbol{T}_t$, $\boldsymbol{H}_t$, are functionally related to a vector of hyperparameters, $\boldsymbol{\theta}$. When $\boldsymbol{a}_1$

and $\boldsymbol{P}_1$ are known and finite (as when the system matrices are time-invariant and $\boldsymbol{y}_t$ is stationary) we shall refer to (1) as a *standard* state space model.

The Kalman filter (KF) is a well-known recursive algorithm for computing the minimum mean square estimator of $\boldsymbol{\alpha}_t$ and its mean square error (MSE) matrix conditional on $\boldsymbol{Y}_{t-1} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{t-1}\}$. Defining

$$\boldsymbol{a}_t = \mathrm{E}(\boldsymbol{\alpha}_t|\boldsymbol{Y}_{t-1}), \;\; \mathrm{MSE}(\boldsymbol{a}_t) = \sigma^2\boldsymbol{P}_t = \mathrm{E}[(\boldsymbol{\alpha}_t - \boldsymbol{a}_t)(\boldsymbol{\alpha}_t - \boldsymbol{a}_t)'|\boldsymbol{Y}_{t-1}],$$

the filter consists of the following recursions (Anderson and Moore, 1979, sec. 5.4):

$$\begin{array}{rclrcl} \boldsymbol{\nu}_t & = & \boldsymbol{y}_t - \boldsymbol{Z}_t\boldsymbol{a}_t, & \boldsymbol{F}_t & = & \boldsymbol{Z}_t\boldsymbol{P}_t\boldsymbol{Z}_t' + \boldsymbol{G}_t\boldsymbol{G}_t' \\ q_t & = & q_{t-1} + \boldsymbol{\nu}_t'\boldsymbol{F}_t^{-1}\boldsymbol{\nu}_t, & \boldsymbol{K}_t & = & (\boldsymbol{T}_t\boldsymbol{P}_t\boldsymbol{Z}_t' + \boldsymbol{H}_t\boldsymbol{G}_t')\boldsymbol{F}_t^{-1} \\ \boldsymbol{a}_{t+1} & = & \boldsymbol{T}_t\boldsymbol{a}_t + \boldsymbol{K}_t\boldsymbol{\nu}_t, & \boldsymbol{P}_{t+1} & = & \boldsymbol{T}_t\boldsymbol{P}_t\boldsymbol{T}_t' + \boldsymbol{H}_t\boldsymbol{H}_t' - \boldsymbol{K}_t\boldsymbol{F}_t\boldsymbol{K}_t' \end{array} \tag{2}$$

with $q_0 = 0$; $\boldsymbol{\nu}_t = \boldsymbol{y}_t - \mathrm{E}(\boldsymbol{y}_t|\boldsymbol{Y}_{t-1})$ are the filter innovations or one-step-ahead prediction errors, with MSE matrix $\sigma^2\boldsymbol{F}_t$. The log-likelihood for the model is, apart from a constant term,

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{2}\left[NT\ln\sigma^2 + \sum_{t=1}^{T}\ln|\boldsymbol{F}_t| + \sigma^{-2}q_T\right],$$

where $q_T = \sum_{t=1}^{T}\boldsymbol{\nu}_t'\boldsymbol{F}_t^{-1}\boldsymbol{\nu}_t$, and the maximum likelihood estimate of σ^2 is $\hat{\sigma}^2 = q_T/NT$.

Diagnostic checking is usually carried out using the standardised innovations $\boldsymbol{F}_t^{-1/2}\boldsymbol{\nu}_t \sim \mathrm{NID}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$, which play a role in detecting various types of misspecifications, such as serial correlation, heteroscedasticity, nonnormality and structural change (CUSUM).

Other sets of residuals are built upon the output of the smoothing filter (De Jong, 1988, 1989, Kohn and Ansley, 1989):

$$\begin{array}{rclrcl} \boldsymbol{u}_t & = & \boldsymbol{F}_t^{-1}\boldsymbol{\nu}_t - \boldsymbol{K}_t'\boldsymbol{r}_t, & \boldsymbol{M}_t & = & \boldsymbol{F}_t^{-1} + \boldsymbol{K}_t'\boldsymbol{N}_t\boldsymbol{K}_t, \\ \boldsymbol{r}_{t-1} & = & \boldsymbol{Z}_t'\boldsymbol{F}_t^{-1}\boldsymbol{\nu}_t + \boldsymbol{L}_t'\boldsymbol{r}_t, & \boldsymbol{N}_{t-1} & = & \boldsymbol{Z}_t'\boldsymbol{F}_t^{-1}\boldsymbol{Z}_t + \boldsymbol{L}_t'\boldsymbol{N}_t\boldsymbol{L}_t, \end{array} \tag{3}$$

$\boldsymbol{L}_t = \boldsymbol{T}_t - \boldsymbol{K}_t\boldsymbol{Z}_t$, started with $\boldsymbol{r}_T = \boldsymbol{0}$ and $\boldsymbol{N}_T = \boldsymbol{0}$; $\boldsymbol{u}_t$ is sometimes termed a *smoothing error* (Harvey et *al.*, 1998).

Auxiliary residuals (Koopman, 1993) are the smoothed estimators of disturbances associated with the unobserved components and are based on the disturbance smoother $\mathrm{E}(\boldsymbol{\varepsilon}_t|Y_T) = \boldsymbol{G}_t'\boldsymbol{u}_t + \boldsymbol{H}_t'\boldsymbol{r}_t$. Once they are standardised by their unconditional standard deviation, they provide test statistics for outliers and structural change in the state components (Harvey and Koopman, 1992, De Jong and Penzer, 1998). Unlike the standardised innovations, the auxiliary residuals are serially correlated; Harvey and Koopman (1992) derive their autocorrelation structure and show how they can be employed to form appropriate tests of normality. When the measurement and the transition equation disturbances are uncorrelated, i.e. $\boldsymbol{G}_t\boldsymbol{H}_t' = \boldsymbol{0}$, the irregular auxiliary residual

$$\boldsymbol{y}_t - \boldsymbol{Z}_t\mathrm{E}(\boldsymbol{\alpha}_t|\boldsymbol{Y}_T) = \boldsymbol{G}_t\mathrm{E}(\boldsymbol{\varepsilon}_t|\boldsymbol{Y}_T) = \boldsymbol{G}_t\boldsymbol{G}_t'\boldsymbol{u}_t,$$

standardised by the estimated unconditional covariance matrix, $\hat{\sigma}^2(\boldsymbol{G}_t\boldsymbol{G}_t'\boldsymbol{M}_t\boldsymbol{G}_t\boldsymbol{G}_t')$, corresponds to what is known in the regression literature as an internally studentised residuals (Kohn and Ansley, 1989).

The residual arising from deletion of the observation at time $t = i$ (prediction or deletion residual) is (De Jong, 1988, Kohn and Ansley, 1989):

$$\boldsymbol{y}_i - \mathrm{E}(\boldsymbol{y}_i|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-1}, \boldsymbol{y}_{i+1}, \ldots, \boldsymbol{y}_T) = \boldsymbol{M}_i^{-1}\boldsymbol{u}_i$$

with unconditional covariance $\sigma^2\boldsymbol{M}_i^{-1}$. The estimate of σ^2 with $\boldsymbol{y}_i$ deleted is

$$\hat{\sigma}^2_{(i)} = \frac{q_T - \boldsymbol{u}_i'\boldsymbol{M}_i^{-1}\boldsymbol{u}_i}{N(T-1)}$$

and is related to $\hat{\sigma}^2$ by $N(T-1)\hat{\sigma}^2_{(i)} = NT\hat{\sigma}^2 - \boldsymbol{u}_i'\boldsymbol{M}_i^{-1}\boldsymbol{u}_i$, which mirrors analogous computations carried out in the linear regression case (Cook and Weisberg, 1982, p. 20, Atkinson, 1985, p. 21). The vector of cross-validatory or externally studentised residuals at time i can then be defined, and under normality the quadratic form $\hat{\tau}_{(i)} = \hat{\sigma}_{(i)}^{-2}\boldsymbol{u}_i'\boldsymbol{M}_i^{-1}\boldsymbol{u}_i$ is F-distributed with 1 and $N(T-1)$ degrees of freedom and can be used to test if the i-th observation is outlying. This result stems from the independence of $\boldsymbol{u}_i'\boldsymbol{M}_i^{-1}\boldsymbol{u}_i$ from $\hat{\sigma}_{(i)}^2$.

3. Leave-k-out Diagnostics for Standard State Space Models

We now address the issue of leaving out k consecutive observations: the virtues of this strategy for detecting patches of outliers in the ARIMA framework have been advocated by Bruce and Martin (1989). Assuming that observations $\boldsymbol{y}_{i-k+1}, \ldots, \boldsymbol{y}_i$ are deleted and denoting $\boldsymbol{y}_{(I)}$ the stack of the deleted observations, Kohn and Ansley (1989) showed that the vector of deletion residuals is

$$\boldsymbol{y}_{(I)} - \mathrm{E}(\boldsymbol{y}_{(I)}|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-k}, \boldsymbol{y}_{i+1}, \ldots, \boldsymbol{y}_T) = \boldsymbol{M}_{(I)}^{-1}\boldsymbol{u}_{(I)} \tag{4}$$

The matrix $\boldsymbol{M}_{(I)}$ has dimension $Nk \times Nk$, with diagonal blocks $\boldsymbol{M}_t, t = i-k+1, \ldots, i$ and off-diagonal blocks $\boldsymbol{M}_{ts} = \sigma^{-2}\mathrm{Cov}(\boldsymbol{u}_t, \boldsymbol{u}_s)$; also, $\boldsymbol{u}_{(I)} = [\boldsymbol{u}_{i-k+1}', \ldots, \boldsymbol{u}_i']'$. The deletion residuals have unconditional covariance matrix $\sigma^2\boldsymbol{M}_{(I)}^{-1}$.

The statistic

$$\hat{\tau}_{(I)} = \frac{\boldsymbol{u}_{(I)}'\boldsymbol{M}_{(I)}^{-1}\boldsymbol{u}_{(I)}}{k\hat{\sigma}_{(I)}^2}, \tag{5}$$

with

$$\hat{\sigma}_{(I)}^2 = \frac{q_T - \boldsymbol{u}_{(I)}'\boldsymbol{M}_{(I)}^{-1}\boldsymbol{u}_{(I)}}{N(T-k)},$$

provides a test that the observations are jointly outlying. Under normality the exact distribution of $\hat{\tau}_{(I)}$ is $\mathrm{F}(k, N(T-k))$.

As far as the computation of $\hat{\tau}_{(I)}$ is concerned, note that the off-diagonal blocks of $\boldsymbol{M}_{(I)}$ are not produced automatically by the KF and the smoothing filter for the complete observations, although explicit formulae are given in De Jong (1989) and Kohn and Ansley (1989). Furthermore, for multivariate state space model and/or large k, direct inversion of the matrix $\boldsymbol{M}_{(I)}$ may not be computationally attractive.

The alternative strategy proposed here is to consider a transformation of the smoothing errors, $\boldsymbol{u}_{(I)}$, which diagonalises the matrix in question. This is achieved by running backwards the Kalman filter on the pseudo-model made up of the measurement equation $\boldsymbol{u}_t = \boldsymbol{F}_t^{-1}\boldsymbol{\nu}_t - \boldsymbol{K}_t'\boldsymbol{r}_t$ and transition equation $\boldsymbol{r}_{t-1} = \boldsymbol{Z}_t'\boldsymbol{F}_t^{-1}\boldsymbol{\nu}_t + \boldsymbol{L}_t'\boldsymbol{r}_t$, where $\boldsymbol{F}_t^{-1/2}\boldsymbol{\nu}_t$ act as disturbances, $\boldsymbol{u}_t$ are the observations and $\boldsymbol{r}_t$ the states.

This amounts to running the filter

$$
\begin{array}{rclrcl}
\boldsymbol{u}_t^* & = & \boldsymbol{u}_t + \boldsymbol{K}_t'\boldsymbol{r}_t^*, & \boldsymbol{M}_t^* & = & \boldsymbol{F}_t^{-1} + \boldsymbol{K}_t'\boldsymbol{N}_t^*\boldsymbol{K}_t, \\
q_{t-1}^* & = & q_t^* + \boldsymbol{u}_t^{*'}\boldsymbol{M}_t^{*-1}\boldsymbol{u}_t^*, & \boldsymbol{K}_t^* & = & (\boldsymbol{Z}_t'\boldsymbol{F}_t^{-1} - \boldsymbol{L}_t'\boldsymbol{N}_t^*\boldsymbol{K}_t)\boldsymbol{M}_t^{*-1}, \\
\boldsymbol{r}_{t-1}^* & = & \boldsymbol{L}_t'\boldsymbol{r}_t^* + \boldsymbol{K}_t^*\boldsymbol{u}_t^*, & \boldsymbol{N}_{t-1}^* & = & \boldsymbol{Z}_t'\boldsymbol{F}_t^{-1}\boldsymbol{Z}_t + \boldsymbol{L}_t'\boldsymbol{N}_t^*\boldsymbol{L}_t - \boldsymbol{K}_t^*\boldsymbol{M}_t^*\boldsymbol{K}_t^{*'},
\end{array} \tag{6}
$$

for $t = i, i-1, \ldots, i-k+1$. The filter is initialised by the unconditional mean and covariance matrix of $\boldsymbol{r}_i$, that is $\boldsymbol{r}_i^* = \boldsymbol{0}$ and $\boldsymbol{N}_i^* = \boldsymbol{N}_i$, and by $q_i^* = 0$.

The filter (6) performs the Choleski block triangular factorisation $\boldsymbol{M}_{(I)} = \boldsymbol{\mathcal{C}}^{-1}\boldsymbol{M}_{(I)}^*\boldsymbol{\mathcal{C}}^{-1\prime}$, where $\boldsymbol{M}_{(I)}^* = \mathrm{diag}(\boldsymbol{M}_{i-k+1}^*, \ldots, \boldsymbol{M}_i^*)$ and $\boldsymbol{\mathcal{C}}$ is an upper triangular matrix with identity blocks on the main diagonal, so that $|\boldsymbol{\mathcal{C}}| = 1$, and $\boldsymbol{u}_{(I)}^* = [\boldsymbol{u}_{i-k+1}^{*'}, \ldots, \boldsymbol{u}_i^{*'}]' = \boldsymbol{\mathcal{C}}\boldsymbol{u}_{(I)}$. This allows us to write

$$
\hat{\tau}_{(I)} = \frac{q_{i-k+1}^*}{q_T - q_{i-k+1}^*}\frac{N(T-k)}{k}. \tag{7}
$$

The output of the filter is the set of uncorrelated quantities $\boldsymbol{u}_t^*$ with unconditional covariance $\sigma^2 \boldsymbol{M}_t^*$, such that:

$$\boldsymbol{M}_{i-j}^{*-1}\boldsymbol{u}_{i-j}^* = \boldsymbol{y}_{i-j} - \mathrm{E}(\boldsymbol{y}_{i-j}|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-j-1}, \boldsymbol{y}_{i+1}, \ldots, \boldsymbol{y}_T), \;\; j = 0, 1, \ldots, k-1.$$

Therefore, once this filter is run for the maximum k desired, leave-r-out diagnostics for $r < k$ are immediately available. Note that, when applied for $i = T$, it produces $\boldsymbol{M}_t^{*-1}\boldsymbol{u}_t^* = \boldsymbol{\nu}_t$, $(T - k + 1) \leq t \leq T$.

4. Nonstationary State Space Models with Regression Effects

When regression effects and nonstationary state components are present, the general state space form for the complete observations is formulated as follows:

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{Z}_t\boldsymbol{\alpha}_t + \boldsymbol{X}_t\boldsymbol{\delta} + \boldsymbol{G}_t\boldsymbol{\varepsilon}_t, \qquad t = 1, 2, \ldots, T, \\ \boldsymbol{\alpha}_{t+1} &= \boldsymbol{T}_t\boldsymbol{\alpha}_t + \boldsymbol{W}_t\boldsymbol{\delta} + \boldsymbol{H}_t\boldsymbol{\varepsilon}_t, \end{aligned} \tag{8}$$

with $\boldsymbol{\alpha}_1 = \boldsymbol{W}_0\boldsymbol{\gamma} + \boldsymbol{H}_0\boldsymbol{\varepsilon}_0$. The matrices $\boldsymbol{X}_t$ and $\boldsymbol{W}_t$ contain the regression effects and have column dimension k. The vector $\boldsymbol{\beta} = [\boldsymbol{\gamma}', \boldsymbol{\delta}']'$ is assumed diffuse, i.e. $\boldsymbol{\beta} \sim \boldsymbol{N}(\boldsymbol{0}, \kappa\boldsymbol{I})$, $\kappa \to \infty$, and of dimension $d + k$, where d is the number of nonstationary components of $\boldsymbol{\alpha}_t$; $\boldsymbol{W}_0$ is a selection matrix associating $\boldsymbol{\gamma}$ to the nonstationary components in the state vector.

The statistical treatment of model (8) entails augmenting the KF (2) by the following recursions (De Jong, 1991):

$$\begin{aligned} \boldsymbol{V}_t &= [\boldsymbol{0}, \boldsymbol{X}_t] - \boldsymbol{Z}_t\boldsymbol{A}_t, & \boldsymbol{A}_{t+1} &= \boldsymbol{T}_t\boldsymbol{A}_t - [\boldsymbol{0}, \boldsymbol{W}_t] + \boldsymbol{K}_t\boldsymbol{V}_t, \\ \boldsymbol{S}_t &= \boldsymbol{S}_{t-1} + \boldsymbol{V}_t'\boldsymbol{F}_t^{-1}\boldsymbol{V}_t, & \boldsymbol{s}_t &= \boldsymbol{s}_{t-1} + \boldsymbol{V}_t'\boldsymbol{F}_t^{-1}\boldsymbol{\nu}_t, \end{aligned} \tag{9}$$

with starting conditions: $\boldsymbol{A}_1 = [\boldsymbol{W}_0, \boldsymbol{0}]$, $\boldsymbol{S}_0 = \boldsymbol{0}$ and $\boldsymbol{s}_0 = \boldsymbol{0}$, where the column dimension of $\boldsymbol{V}_t$ and $\boldsymbol{A}_t$ is $d + k$. This amounts to running the

KF on the d columns of zeros and on $\boldsymbol{X}_t$ and accumulating $\boldsymbol{V}_t'\boldsymbol{F}_t^{-1}\boldsymbol{V}_t$ and $\boldsymbol{V}_t'\boldsymbol{F}_t^{-1}\boldsymbol{\nu}_t$.

De Jong (1991) shows that as $\kappa \to \infty$ the limiting expression for $\mathrm{E}(\boldsymbol{\beta}|\boldsymbol{Y}_T)$ and $\mathrm{MSE}(\boldsymbol{\beta}|\boldsymbol{Y}_T)$ are respectively $\boldsymbol{S}_T^{-1}\boldsymbol{s}_T$, and $\sigma^2\boldsymbol{S}_T^{-1}$, and that it is possible to define a proper likelihood, taking the form:

$$\mathcal{L}_\infty(\boldsymbol{\theta}) = -\frac{1}{2}\left[NT^* \ln \sigma^2 + \sum_{t=1}^{T} \ln |\boldsymbol{F}_t| + \ln |\boldsymbol{S}_T| + \sigma^{-2}(q_T - \boldsymbol{s}_T'\boldsymbol{S}_T^{-1}\boldsymbol{s}_T)\right],$$

with $T^* = T - d - k$. The maximum likelihood estimate of σ^2 is $\hat{\sigma}^2 = (q_T - \boldsymbol{s}_T'\boldsymbol{S}_T^{-1}\boldsymbol{s}_T)/NT^*$. The notation $\mathcal{L}_\infty(\boldsymbol{\theta})$ stresses that this is a diffuse log-likelihood, based on a rank T^* transformation of the observations with unit Jacobian, which makes the data invariant to $\boldsymbol{\beta}$.

Diagnostic checking can be performed on the *generalised recursive residuals*, $\hat{\boldsymbol{\nu}}_t = \boldsymbol{\nu}_t - \boldsymbol{V}_t\boldsymbol{S}_{t-1}^{-1}\boldsymbol{s}_{t-1}$, which are uncorrelated with covariance matrix $\sigma^2\hat{\boldsymbol{F}}_t = \sigma^2(\boldsymbol{F}_t + \boldsymbol{V}_t\boldsymbol{S}_{t-1}^{-1}\boldsymbol{V}_t')$. The auxiliary and the deletion residuals under diffuse effects are defined in terms of the output of the smoother (3) augmented by the recursions:

$$\boldsymbol{U}_t = \boldsymbol{F}_t^{-1}\boldsymbol{V}_t - \boldsymbol{K}_t'\boldsymbol{R}_t, \qquad \boldsymbol{R}_{t-1} = \boldsymbol{Z}_t'\boldsymbol{F}_t^{-1}\boldsymbol{V}_t + \boldsymbol{L}_t'\boldsymbol{R}_t,$$

with $\boldsymbol{R}_T = \boldsymbol{0}$.

As far the residual arising when the i-th observation is deleted, De Jong (1988) shows that

$$\boldsymbol{y}_i - \mathrm{E}(\boldsymbol{y}_i|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-1}, \boldsymbol{y}_{i+1}, \ldots, \boldsymbol{y}_T) = \hat{\boldsymbol{M}}_i^{-1}\hat{\boldsymbol{u}}_i$$

where $\hat{\boldsymbol{u}}_i = \boldsymbol{u}_i - \boldsymbol{U}_i\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{M}}_i = \boldsymbol{M}_i - \boldsymbol{U}_i\boldsymbol{S}_T^{-1}\boldsymbol{U}_i'$. The estimate of σ^2 with $\boldsymbol{y}_i$ deleted is $\hat{\sigma}^2_{(i)} = (q_T - \boldsymbol{s}_T'\boldsymbol{S}_T^{-1}\boldsymbol{s}_T - \hat{\boldsymbol{u}}_i\hat{\boldsymbol{M}}_i^{-1}\hat{\boldsymbol{u}}_i)/(N(T^* - k))$.

Interest lies also in assessing how influential is $\boldsymbol{y}_i$ for inferences on $\boldsymbol{\beta}$. Denoting $\hat{\boldsymbol{\beta}}_{(i)}$ the estimate of $\boldsymbol{\beta}$ arising when $\boldsymbol{y}_i$ is dropped, and elaborating results from De Jong and Penzer (1998), the change in the estimate of the initial and regression effects is

$$\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)} = \boldsymbol{S}_T^{-1}\boldsymbol{U}_i'\hat{\boldsymbol{M}}_i^{-1}\hat{\boldsymbol{u}}_i$$

and

$$\sigma^{-2}[\mathrm{MSE}(\hat{\boldsymbol{\beta}}) - \mathrm{MSE}(\hat{\boldsymbol{\beta}}_{(i)})] = -\boldsymbol{S}_T^{-1}\boldsymbol{U}_i'\hat{\boldsymbol{M}}_i^{-1}\boldsymbol{U}_i\boldsymbol{S}_T^{-1}$$

These results can be used to compute influence measures such as the standard and the modified Cook's Distance (Cook and Weisberg, 1982, Atkinson, 1985).

5. Leave-k-out Diagnostics for Nonstationary Models

Let us define

$$\hat{\boldsymbol{u}}_{(I)} = \boldsymbol{u}_{(I)} - \boldsymbol{U}_{(I)}\hat{\boldsymbol{\beta}}, \quad \hat{\boldsymbol{M}}_{(I)} = \boldsymbol{M}_{(I)} - \boldsymbol{U}_{(I)}\boldsymbol{S}_T^{-1}\boldsymbol{U}_{(I)}',$$

where $\hat{\boldsymbol{u}}_{(I)} = [\hat{\boldsymbol{u}}_{i-k+1}', \ldots, \hat{\boldsymbol{u}}_i']'$ and $\boldsymbol{U}_{(I)} = [\boldsymbol{U}_{i-k+1}', \ldots, \boldsymbol{U}_i']'$; the matrix $\hat{\boldsymbol{M}}_{(I)}$ has diagonal blocks $\hat{\boldsymbol{M}}_t, t = i-k+1, \ldots, i$, and off-diagonal blocks $\hat{\boldsymbol{M}}_{ts} = \boldsymbol{M}_{ts} - \boldsymbol{U}_t\boldsymbol{S}_T^{-1}\boldsymbol{U}_s'$.

The vector of deletion residuals can be written

$$\boldsymbol{y}_{(I)} - \mathrm{E}(\boldsymbol{y}_{(I)}|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-k}, \boldsymbol{y}_{i+1}, \ldots, \boldsymbol{y}_T) = \hat{\boldsymbol{M}}_{(I)}^{-1}\hat{\boldsymbol{u}}_{(I)}$$

with unconditional covariance matrix $\sigma^2\hat{\boldsymbol{M}}_{(I)}^{-1}$. The statistic

$$\hat{\tau}_{(I)} = \frac{\hat{\boldsymbol{u}}_{(I)}'\hat{\boldsymbol{M}}_{(I)}^{-1}\hat{\boldsymbol{u}}_{(I)}}{k\hat{\sigma}_{(I)}^2} \tag{10}$$

with

$$\hat{\sigma}_{(I)}^2 = \frac{q_T - \boldsymbol{s}_T'\boldsymbol{S}_T^{-1}\boldsymbol{s}_T - \hat{\boldsymbol{u}}_{(I)}'\hat{\boldsymbol{M}}_{(I)}^{-1}\hat{\boldsymbol{u}}_{(I)}}{N(T^* - k)}$$

provides a test that $\boldsymbol{y}_{(I)}$ is a multiple outlier; its reference distribution is $\mathrm{F}(k, N(T^* - k))$.

The change in the estimate of the initial and regression effects when the set (I) is deleted

$$\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(I)} = \boldsymbol{S}_T^{-1}\boldsymbol{U}'_{(I)}\hat{\boldsymbol{M}}_{(I)}^{-1}\hat{\boldsymbol{u}}_{(I)} \tag{11}$$

and $\mathrm{MSE}(\hat{\boldsymbol{\beta}}) - \mathrm{MSE}(\hat{\boldsymbol{\beta}}_{(I)}) = -\boldsymbol{S}_T^{-1}\boldsymbol{U}'_{(I)}\hat{\boldsymbol{M}}_{(I)}^{-1}\boldsymbol{U}_{(I)}\boldsymbol{S}_T^{-1}$.

All quantities that are relevant for computation of leave-k-out diagnostics are made available by the filter (6), augmented by the recursions

$$\begin{array}{rclrcl}
\boldsymbol{U}_t^* & = & \boldsymbol{U}_t + \boldsymbol{K}_t'\boldsymbol{R}_t^*, & \boldsymbol{R}_{t-1}^* & = & \boldsymbol{L}_t'\boldsymbol{R}_t^* + \boldsymbol{K}_t^*\boldsymbol{U}_t^*, \\
\hat{\boldsymbol{u}}_t^* & = & \boldsymbol{u}_t^* - \boldsymbol{U}_t^*\boldsymbol{b}_t^*, & \hat{\boldsymbol{M}}_t^* & = & \boldsymbol{M}_t^* - \boldsymbol{U}_t^*\boldsymbol{B}_t^*\boldsymbol{U}_t^{*\prime} \\
\tilde{\boldsymbol{K}}_t & = & \boldsymbol{B}_t^*\boldsymbol{U}_t^{*\prime}\hat{\boldsymbol{M}}_t^{*-1}, & & & \\
\boldsymbol{b}_{t-1}^* & = & \boldsymbol{b}_t^* - \tilde{\boldsymbol{K}}_t\hat{\boldsymbol{u}}_t^*, & \boldsymbol{B}_{t-1}^* & = & \boldsymbol{B}_t^* + \tilde{\boldsymbol{K}}_t\hat{\boldsymbol{M}}_t^*\tilde{\boldsymbol{K}}_t'
\end{array} \tag{12}$$

for $t = i, i-1, \ldots, i-k+1$, and initial values $\boldsymbol{b}_i^* = \hat{\boldsymbol{\beta}}, \boldsymbol{B}_i^* = \boldsymbol{S}_T^{-1}$ and $\boldsymbol{R}_i^* = \boldsymbol{0}$. These computations parallel the augmented KF (9) and the formulae for the generalised recursive residuals. The quantities $\boldsymbol{b}_t^*$ and $\boldsymbol{B}_t^*$ keep track of the changes in the estimate of $\boldsymbol{\beta}$ and its MSE matrix arising from sequential deletion of observations, as will be shown below. The augmented filter delivers the set of uncorrelated quantities $\hat{\boldsymbol{u}}_t^*$ with unconditional covariance $\sigma^2\hat{\boldsymbol{M}}_t^*$, such that:

$$\hat{\boldsymbol{M}}_{i-j}^{*-1}\hat{\boldsymbol{u}}_{i-j}^* = \boldsymbol{y}_{i-j} - \mathrm{E}(\boldsymbol{y}_{i-j}|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-j-1}, \boldsymbol{y}_{i+1}, \ldots, \boldsymbol{y}_T), \quad j = 0, 1, \ldots, k-1.$$

Thus, once this filter is run for the maximum k desired, leave-r-out diagnostics for $r < k$ are immediately available. Also, when applied for $i = T$, it produces $\hat{\boldsymbol{M}}_t^{*-1}\hat{\boldsymbol{u}}_t^* = \hat{\boldsymbol{\nu}}_t$, i.e. the generalised recursive residuals for $(T - k + 1) \leq t \leq T$.

Theorem. The output of the augmented filter consisting of equations (6) and

(12), $\hat{\boldsymbol{u}}_t^*$ and $\hat{\boldsymbol{M}}_t^*$, is used to compute

$$\hat{\boldsymbol{u}}_{(I)}'\hat{\boldsymbol{M}}_{(I)}^{-1}\hat{\boldsymbol{u}}_{(I)} = \sum_{t=i-k+1}^{i} \hat{\boldsymbol{u}}_t^{*'}\hat{\boldsymbol{M}}_t^{*-1}\hat{\boldsymbol{u}}_t^*.$$

Moreover,

$$\boldsymbol{b}_{i-k}^* = \hat{\boldsymbol{\beta}}_{(I)}, \quad \boldsymbol{B}_{i-k}^* = \sigma^{-2}\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{(I)}).$$

Proof. Define $\boldsymbol{B}_{(I)}^* = \mathrm{diag}(\boldsymbol{B}_{i-k+1}^*, \ldots, \boldsymbol{B}_i^*)$, $\boldsymbol{\mathcal{U}}_{(I)}^* = \mathrm{diag}(\boldsymbol{U}_{i-k+1}^*, \ldots, \boldsymbol{U}_i^*)$, $\hat{\boldsymbol{M}}_{(I)}^* = \mathrm{diag}(\hat{\boldsymbol{M}}_{i-k+1}^*, \ldots, \hat{\boldsymbol{M}}_i^*)$, $\hat{\boldsymbol{u}}_{(I)}^* = [\hat{\boldsymbol{u}}_{i-k+1}^{*'}, \ldots, \hat{\boldsymbol{u}}_i^{*'}]'$, $\boldsymbol{b}_{(I)}^* = [\boldsymbol{b}_{i-k+1}^{*'}, \ldots, \boldsymbol{b}_i^{*'}]'$,

$$\boldsymbol{\mathcal{D}} = \begin{bmatrix} \mathbf{0} & \boldsymbol{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}, \quad \tilde{\boldsymbol{K}} = \begin{bmatrix} \mathbf{0} & \tilde{\boldsymbol{K}}_{i-k+2} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \tilde{\boldsymbol{K}}_{i-k+3} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \tilde{\boldsymbol{K}}_i \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}.$$

Then, $\tilde{\boldsymbol{K}} = \boldsymbol{\mathcal{D}}\boldsymbol{B}_{(I)}^*\boldsymbol{\mathcal{U}}_{(I)}^{*'}\hat{\boldsymbol{M}}_{(I)}^{*-1}$, and letting $\tilde{\boldsymbol{\mathcal{K}}} = (\boldsymbol{I} - \boldsymbol{\mathcal{D}})^{-1}\tilde{\boldsymbol{K}}$,

$$\boldsymbol{b}_{(I)}^* = (\boldsymbol{i}_k \otimes \hat{\boldsymbol{\beta}}) - \tilde{\boldsymbol{\mathcal{K}}}\hat{\boldsymbol{u}}_{(I)}^*,$$

where $\boldsymbol{i}_k$ is a $k \times 1$ vector of unit elements. Now, recalling that $\boldsymbol{u}_{(I)}^* = \boldsymbol{\mathcal{C}}\boldsymbol{u}_{(I)}$, $\boldsymbol{U}_{(I)}^* = [\boldsymbol{U}_{i-k+1}^{*'}, \ldots, \boldsymbol{U}_i^{*'}]' = \boldsymbol{\mathcal{C}}\boldsymbol{U}_{(I)}$, where the last statement follows from the fact that the same filter is applied to $\boldsymbol{U}_t$, and replacing for $(\boldsymbol{i}_k \otimes \hat{\boldsymbol{\beta}})$ into $\hat{\boldsymbol{u}}_{(I)} = \boldsymbol{\mathcal{C}}^{-1}[\boldsymbol{u}_{(I)}^* - \boldsymbol{\mathcal{U}}_{(I)}^*(\boldsymbol{i}_k \otimes \hat{\boldsymbol{\beta}})]$, we obtain

$$\hat{\boldsymbol{u}}_{(I)} = \boldsymbol{\mathcal{C}}^{-1}[\boldsymbol{I} - \boldsymbol{\mathcal{U}}_{(I)}^*\tilde{\boldsymbol{\mathcal{K}}}]\hat{\boldsymbol{u}}_{(I)}^*.$$

In matrix notation, the recursions for $\boldsymbol{B}_t^*$ can be written

$$\boldsymbol{B}_{(I)}^* - \boldsymbol{\mathcal{D}}\boldsymbol{B}_{(I)}^*\boldsymbol{\mathcal{D}}' = \boldsymbol{\mathcal{E}}_k\boldsymbol{S}_T^{-1}\boldsymbol{\mathcal{E}}_k' + \tilde{\boldsymbol{K}}\hat{\boldsymbol{M}}_{(I)}^*\tilde{\boldsymbol{K}}',$$

with $\boldsymbol{\mathcal{E}}_k' = [\mathbf{0}, \cdots, \mathbf{0}, \boldsymbol{I}]$, so that premultiplying for $(\boldsymbol{I} - \boldsymbol{\mathcal{D}})^{-1}$ and postmultiplying for $(\boldsymbol{I} - \boldsymbol{\mathcal{D}})^{-1\prime}$ and rearranging, we get

$$\boldsymbol{B}_{(I)}^* + (\boldsymbol{I} - \boldsymbol{\mathcal{D}})^{-1}\boldsymbol{\mathcal{D}}\boldsymbol{B}_{(I)}^* + \boldsymbol{B}_{(I)}^*\boldsymbol{\mathcal{D}}'(\boldsymbol{I} - \boldsymbol{\mathcal{D}})^{-1\prime} = (\boldsymbol{i}_k\boldsymbol{i}_k' \otimes \boldsymbol{S}_T^{-1}) + \tilde{\boldsymbol{\mathcal{K}}}\hat{\boldsymbol{M}}_{(I)}^*\tilde{\boldsymbol{\mathcal{K}}}'.$$

This result is used to provide an alternative expression for $\hat{\boldsymbol{M}}_{(I)}$:

$$\begin{aligned}
\hat{\boldsymbol{M}}_{(I)} &= \boldsymbol{M}_{(I)} - \boldsymbol{U}_{(I)}\boldsymbol{S}_T^{-1}\boldsymbol{U}'_{(I)} \\
&= \boldsymbol{\mathcal{C}}^{-1}[\boldsymbol{M}^*_{(I)} - \boldsymbol{\mathcal{U}}^*_{(I)}(\boldsymbol{i}_k\boldsymbol{i}'_k \otimes \boldsymbol{S}_T^{-1})\boldsymbol{\mathcal{U}}^{*'}_{(I)}]\boldsymbol{\mathcal{C}}^{-1\prime} \\
&= \boldsymbol{\mathcal{C}}^{-1}[\boldsymbol{M}^*_{(I)} - \boldsymbol{\mathcal{U}}^*_{(I)}\boldsymbol{B}^*_{(I)}\boldsymbol{\mathcal{U}}^{*'}_{(I)} - \boldsymbol{\mathcal{U}}^*_{(I)}(\boldsymbol{I}-\boldsymbol{\mathcal{D}})^{-1}\boldsymbol{\mathcal{D}}\boldsymbol{B}^*_{(I)}\boldsymbol{\mathcal{U}}^{*'}_{(I)} - \\
&\quad \boldsymbol{\mathcal{U}}^*_{(I)}\boldsymbol{B}^*_{(I)}\boldsymbol{\mathcal{D}}'(\boldsymbol{I}-\boldsymbol{\mathcal{D}})^{-1\prime}\boldsymbol{\mathcal{U}}^{*'}_{(I)} + \boldsymbol{\mathcal{U}}^*_{(I)}\tilde{\boldsymbol{\mathcal{K}}}\hat{\boldsymbol{M}}^*_{(I)}\tilde{\boldsymbol{\mathcal{K}}}'\boldsymbol{\mathcal{U}}^{*'}_{(I)}]\boldsymbol{\mathcal{C}}^{-1\prime} \\
&= \boldsymbol{\mathcal{C}}^{-1}[\boldsymbol{I} - \boldsymbol{\mathcal{U}}^*_{(I)}\tilde{\boldsymbol{\mathcal{K}}}]\hat{\boldsymbol{M}}^*_{(I)}[\boldsymbol{I} - \boldsymbol{\mathcal{U}}^*_{(I)}\tilde{\boldsymbol{\mathcal{K}}}]'\boldsymbol{\mathcal{C}}^{-1\prime}
\end{aligned}$$

where the last expression is derived writing $\boldsymbol{M}^*_{(I)} - \boldsymbol{\mathcal{U}}^*_{(I)}\boldsymbol{B}^*_{(I)}\boldsymbol{\mathcal{U}}^{*'}_{(I)} = \hat{\boldsymbol{M}}^*_{(I)}$ and

$$(\boldsymbol{I}-\boldsymbol{\mathcal{D}})^{-1}\boldsymbol{\mathcal{D}}\boldsymbol{B}^*_{(I)}\boldsymbol{\mathcal{U}}^{*'}_{(I)} = (\boldsymbol{I}-\boldsymbol{\mathcal{D}})^{-1}\tilde{\boldsymbol{K}}\hat{\boldsymbol{M}}^*_{(I)} = \tilde{\boldsymbol{\mathcal{K}}}\hat{\boldsymbol{M}}^*_{(I)}.$$

Hence, $\hat{\boldsymbol{u}}'_{(I)}\hat{\boldsymbol{M}}^{-1}_{(I)}\hat{\boldsymbol{u}}_{(I)} = \hat{\boldsymbol{u}}^{*'}_{(I)}\hat{\boldsymbol{M}}^{*-1}_{(I)}\hat{\boldsymbol{u}}^*_{(I)}$.

Let now $\boldsymbol{\mathcal{E}}'_1 = [\boldsymbol{I}, \boldsymbol{0}, \ldots, \boldsymbol{0}]$. At the end of a run of the filter $\boldsymbol{b}^*_{i-k}$ is generated; this can be written:

$$\begin{aligned}
\boldsymbol{b}^*_{i-k} &= \boldsymbol{\mathcal{E}}'_1\boldsymbol{b}^*_{(I)} - \tilde{\boldsymbol{K}}_{i-k+1}\hat{\boldsymbol{u}}^*_{i-k+1} \\
&= \boldsymbol{\mathcal{E}}'_1\boldsymbol{b}^*_{(I)} - \boldsymbol{B}^*_{i-k+1}\boldsymbol{U}^*_{i-k+1}\hat{\boldsymbol{M}}^{*-1}_{i-k+1}\hat{\boldsymbol{u}}^*_{i-k+1} \\
&= \hat{\boldsymbol{\beta}} - \boldsymbol{\mathcal{E}}'_1\left[\tilde{\boldsymbol{\mathcal{K}}} + \boldsymbol{B}^*_{(I)}\boldsymbol{\mathcal{U}}^*_{(I)}\hat{\boldsymbol{M}}^{*-1}_{(I)}\right]\hat{\boldsymbol{u}}^*_{(I)} \\
&= \hat{\boldsymbol{\beta}} - \boldsymbol{\mathcal{E}}'_1\left[(\boldsymbol{I}-\boldsymbol{\mathcal{D}})^{-1}\boldsymbol{\mathcal{D}}\boldsymbol{B}^*_{(I)} + \boldsymbol{B}^*_{(I)}\right](\boldsymbol{I} - \boldsymbol{\mathcal{U}}^{*'}_{(I)}\tilde{\boldsymbol{\mathcal{K}}}')\boldsymbol{\mathcal{U}}^{*'}_{(I)}\boldsymbol{\mathcal{C}}'^{-1}\hat{\boldsymbol{M}}^{-1}_{(I)}\hat{\boldsymbol{u}}_{(I)} \\
&= \hat{\boldsymbol{\beta}} - \boldsymbol{\mathcal{E}}'_1\left[(\boldsymbol{I}-\boldsymbol{\mathcal{D}})^{-1}\boldsymbol{\mathcal{D}}\boldsymbol{B}^*_{(I)} + \boldsymbol{B}^*_{(I)} - \tilde{\boldsymbol{\mathcal{K}}}\hat{\boldsymbol{M}}^*_{(I)}\tilde{\boldsymbol{\mathcal{K}}}' - \boldsymbol{B}^*_{(I)}\boldsymbol{\mathcal{U}}^{*'}_{(I)}\tilde{\boldsymbol{\mathcal{K}}}'\right]\boldsymbol{\mathcal{U}}^{*'}_{(I)}\boldsymbol{\mathcal{C}}'^{-1}\hat{\boldsymbol{M}}^{-1}_{(I)}\hat{\boldsymbol{u}}_{(I)} \\
&= \hat{\boldsymbol{\beta}} - \boldsymbol{\mathcal{E}}'_1\left[\boldsymbol{B}^*_{(I)}\boldsymbol{\mathcal{D}}'(\boldsymbol{I}-\boldsymbol{\mathcal{D}}')^{-1} - (\boldsymbol{i}_k\boldsymbol{i}'_k \otimes \boldsymbol{S}_T^{-1}) - \boldsymbol{B}^*_{(I)}\boldsymbol{\mathcal{U}}^{*'}_{(I)}\tilde{\boldsymbol{\mathcal{K}}}'\right]\boldsymbol{\mathcal{U}}^{*'}_{(I)}\boldsymbol{\mathcal{C}}'^{-1}\hat{\boldsymbol{M}}^{-1}_{(I)}\hat{\boldsymbol{u}}_{(I)} \\
&= \hat{\boldsymbol{\beta}} + \boldsymbol{S}_T^{-1}\boldsymbol{U}'_{(I)}\hat{\boldsymbol{M}}^{-1}_{(I)}\hat{\boldsymbol{u}}_{(I)}
\end{aligned}$$

as $\boldsymbol{\mathcal{E}}'_1\boldsymbol{B}^*_{(I)}\boldsymbol{\mathcal{D}}'(\boldsymbol{I}-\boldsymbol{\mathcal{D}}')^{-1} = \boldsymbol{0}$ and $\boldsymbol{\mathcal{E}}'_1\boldsymbol{B}^*_{(I)}\boldsymbol{\mathcal{U}}^{*'}_{(I)}\tilde{\boldsymbol{\mathcal{K}}}' = \boldsymbol{0}$. Comparing with (11) gives $\boldsymbol{b}^*_{i-k} = \hat{\boldsymbol{\beta}}_{(I)}$. Finally,

$$\begin{aligned}
\boldsymbol{B}^*_{i-k} &= \boldsymbol{\mathcal{E}}'_1\boldsymbol{B}^*_{(I)}\boldsymbol{\mathcal{E}}_1 + \tilde{\boldsymbol{K}}_{i-k+1}\hat{\boldsymbol{M}}^*_{i-k+1}\tilde{\boldsymbol{K}}'_{i-k+1} \\
&= \boldsymbol{\mathcal{E}}'_1\left[\boldsymbol{B}^*_{(I)} + \boldsymbol{B}^*_{(I)}\boldsymbol{\mathcal{U}}^{*'}_{(I)}\hat{\boldsymbol{M}}^{*-1}_{(I)}\boldsymbol{\mathcal{U}}^*_{(I)}\boldsymbol{B}^*_{(I)}\right]\boldsymbol{\mathcal{E}}_1 \\
&= \boldsymbol{S}_T^{-1} + \boldsymbol{\mathcal{E}}'_1\boldsymbol{B}^*_{(I)}(\boldsymbol{I} - \boldsymbol{\mathcal{U}}^{*'}_{(I)}\tilde{\boldsymbol{\mathcal{K}}}')\boldsymbol{\mathcal{U}}^{*'}_{(I)}\boldsymbol{\mathcal{C}}'^{-1}\hat{\boldsymbol{M}}^{-1}_{(I)}\boldsymbol{\mathcal{C}}^{-1}\boldsymbol{\mathcal{U}}^*_{(I)}(\boldsymbol{I} - \tilde{\boldsymbol{\mathcal{K}}}\boldsymbol{\mathcal{U}}^*_{(I)})\boldsymbol{B}^*_{(I)}\boldsymbol{\mathcal{E}}_1
\end{aligned}$$

gives, after some simple algebra, $\boldsymbol{B}_{i-k} = \boldsymbol{S}_T^{-1} + \boldsymbol{S}_T^{-1}\boldsymbol{U}'_{(I)}\hat{\boldsymbol{M}}^{-1}_{(I)}\boldsymbol{U}_{(I)}\boldsymbol{S}_T^{-1}$. ■

6. Illustration of Leave-k-out Diagnostics for Nonstationary Models

We illustrate the use of leave-k-out diagnostic with reference to the logarithm of the US index of industrial production for the *Textile* sector. This is a quarterly series available for the period 1947.1-1996.4, that has been considered in Proietti (1999). The series, displayed in the first panel of figure 1, can be adequately described by a linear stochastic trend plus trigonometric stochastic seasonality and a nonlinear cyclical component, where the nonlinearity arises as a consequence of the type of asymmetry that has been labelled *steepness* by Sichel (1993): this occurs when troughs are deeper than peaks so that the cyclical dynamics of the series in the vicinity of a trough are, loosely speaking, "faster" and characterised by higher amplitude. In Proietti (1998) a structural model with smooth transition in the damping factor and the frequency of the cycle is fitted, where the transition variable is a filtered estimate of the cycle and the transition mechanism is exponential (see Teräsvirta, 1998).

In this section we fit a linear structural time series model and we demonstrate that leave-k-out diagnostics are useful in detecting patches of observations that are not adequately fitted by the linear model itself. The model we entertain is (8) with system matrices: $\boldsymbol{Z}_t = [1, 0, 1, 0, 1, 1, 0]$, $\boldsymbol{G}_t = \mathbf{0}$, $\boldsymbol{T}_t = \text{diag}(\boldsymbol{T}_\mu, \boldsymbol{T}_\gamma, \boldsymbol{T}_\psi)$, where

$$\boldsymbol{T}_\mu = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{T}_\gamma = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \; \boldsymbol{T}_\psi = \rho \begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix},$$

$\boldsymbol{H}_t = \text{diag}(\sigma_\eta, \sigma_\zeta, \sigma_\omega, \sigma_\omega, \sigma_\omega, \sigma_\kappa, \sigma_\kappa)$, and

$$\boldsymbol{H}_0 = \begin{bmatrix} \mathbf{0} \\ \sigma_\psi \boldsymbol{I}_2 \end{bmatrix}, \quad \boldsymbol{W}_0 = \begin{bmatrix} \boldsymbol{I}_5 \\ \mathbf{0} \end{bmatrix}, \quad \boldsymbol{X}_t = \mathbf{0}, \boldsymbol{W}_t = \mathbf{0}, t > 0,$$

with $\sigma_\kappa^2 = \sigma_\psi^2(1-\rho^2)$. The model represents the stochastic process underlying the series as the sum of a trend component, with transition matrix $\boldsymbol{T}_\mu$ (σ_η^2 and σ_ζ^2 are respectively the variance of the level disturbance and the slope disturbance); a seasonal component, which is the sum of the third and the fifth state components, with transition matrix $\boldsymbol{T}_\gamma$ and common disturbance variance σ_ω^2; a cyclical component, which is modelled by the last two state components, with transition matrix $\boldsymbol{T}_\psi$ and common disturbance variance σ_κ^2; ρ is interpreted as the damping factor, assumed to lie in (0,1), $\lambda \in (0, \pi)$ as the frequency of the cycle, and $\sigma_\psi^2 = \sigma_\kappa^2/(1-\rho^2)$ is the variance of the cycle. Hence there are five nonstationary state components in the model.

Parameter estimation was carried out by maximising the diffuse likelihood and concentrating σ_ψ^2 out of the likelihood function. The resulting parameter estimates are $\hat{\sigma}_\eta^2 = .0001019$, $\hat{\sigma}_\zeta^2 = .0000009$, $\hat{\sigma}_\omega^2 = .0000033$, $\hat{\rho} = .8622$, $\hat{\lambda} = .4966$, corresponding to a period of three years, $\hat{\sigma}_\kappa^2 = .0006697$, and $\hat{\sigma}_\kappa^2 = .0026098$. The evidence for misspecification of the linear model is provided by a the presence of residual kurtosis which makes the Bowman & Shenton normality test highly significant.

A reasonable question is whether this evidence is attributable to the presence of patches of observations that are not adequately fitted by the linear model. In particular we suspect that violation of linearity arises as a consequence of steepness in the vicinity of cyclical troughs. From the inspection of the plot of the series the observations going from 1974.2 to 1975.3 are an obvious candidate.

For the purpose at hand we compute the statistic (10), where the dimension of the set of deleted observations (I) ranges from one to five. The reference distribution is $\mathrm{F}(k, T^* - k)$, with $T^* = 195$.

When k=1 the statistic is a test for the presence of an additive outlier at time i; these "leave-1-out" statistics are displayed in the second panel of figure 1, with the dotted observations representing values significant at the 5% level. It should be noticed that a few isolated points are significant at the beginning of the sample period (1948.2, 1949.1, 1949.2, 1951.2); as a matter of fact, the dynamics at the beginning of the series are somewhat different from the rest of the series. However, the most noticeable fact is that only 1975.1 is flagged as highly outlying; some masking is likely to have taken place, as we suspect that nearby observations should also be outlying; the effect of these is overwhelmed by the effect of the observation 1975.1 and the question is whether joint deletion of consecutive observations can bring to the surface the masked outliers. If this were the case, the strategy of sequentially adding interventions on the basis of leave-1-out diagnostics would not lead very far and would require several iteration to accommodate all outlying effects.

The following panels in figure 1 present leave-k-out diagnostics, for k between 2 and 5. The timing of this diagnostics is such that $\tau_{(I)}$, built on deletion of observations $(i - k + 1, \ldots, i)$, is referred to the midpoint of the deletion interval when k is odd, that is $i - (k - 1)/2$, and to $i - [(k - 1)/2]$ when k is even, where $[x]$ denotes the largest integer less than or equal to x. When two consecutive observations are deleted (panel 3) a patch of four consecutive significant values emerges at the beginning of the series and two consecutive outliers are spotted in 1975: hence, when 1975.1 and 1975.2 are jointly deleted, 1975.2 emerges as outlying. Leave-3-out diagnostic flag four consecutive outliers in the ranging from 1974.4 to 1975.3 and when we move to leave-4-out diagnostics, the third quarter of 1974 enters the set.

It must be recognised that in interpreting these plots a balance has to made between unmasking and the smearing the effect of outliers on adjacent points: for instance as far as leave-5-out diagostics are concerned, it is likely that some smearing has taken place at the beginning of the series (especially in the period 1948-1951); however, the five consecutive outliers, 1974.3-1975.3 are clearly unmasked.

7. Conclusions

The paper has proposed an efficient algorithm based on the Kalman filter run on the auxiliary residuals for computing leave-k-out diagnostics in state space models. The algorithm, in the most general case, that is when diffuse initial and regression effects are present, receives as an input the smoothing errors and their covariance matrix resulting from the augmented smoothing filter and returns a set of whitened errors that are used for computations of the relevant diagnostics. Moreover, the output of the filter enables the assessment of multiple influence on the estimate of the initial and regression effects. The algorithm is also easy to implement, as its recursions mirror the augmented Kalman filter, and it has to be run only for the maximum k required (leave-r-out diagnostics are immediately available).

References

Anderson, B.D.O., and Moore J.B. (1979), *Optimal Filtering*, NJ: Prentice-Hall, Inc., Englewood Cliffs.

Atkinson, A.C. (1985), *Plots, Transformations, and Regression*, Oxford Statistical Science Series, Oxford: Oxford University Press.

Bruce, A.G., and Martin, R.D. (1989), "Leave-k-out diagnostics for time series" (with discussion), *Journal of the Royal Statistical Society*, Series B, 51, 363-424 .

Cook, R.D., and Weisberg, S. (1982), *Residuals and Influence in Regression*, Chapman and Hall.

De Jong, P. (1988), "A Cross-validation Filter for Time Series Models", *Biometrika*, 75, 594-600.

De Jong, P. (1989), "Smoothing and Interpolation with the State Space Model", *Journal of the American Statistical Association*, 84, 1085-1088.

De Jong, P. (1991), "The Diffuse Kalman Filter", *Annals of Statistics*, 19, 1073-1083.

De Jong, P., and Penzer, J. (1998), "Diagnosing Shocks in Time Series", *Journal of the American Statistical Association*, 93 796-806.

Harvey, A.C., and Koopman, S.J. (1992), "Diagnostic Checking of Unobserved Components Time Series Models", *Journal of Business and Economic Statistics*, 10, 377-389.

Harvey, A.C., Koopman, S.J., and Penzer, J. (1998), "Messy Time Series", in *Advances in Econometrics*, eds. T.B. Fomby and R.C. Hill, 13, New York: JAI Press.

Haslett, J., and Hayes, K. (1998), "Residuals for the Linear Model with General Covariance Structure", *Journal of the Royal Statistical Society, B*, 60, 201-215.

Kohn, R., and Ansley, C.F. (1989), "A Fast Algorithm for Signal Extraction, Infuence and Cross-validation in State Space Models", *Biometrika*, 76, 65-79.

Koopman, S.J. (1993), "Disturbance Smoother for State Space Models", *Biometrika*, 80, 117-126.

Proietti, T. (1999), "Characterising Asymmetries in Business Cycles using Smooth Transition Structural Time Series Models", *Studies in Nonlinear Dynamics and Econometrics*, Vol 3.3, 141-156.

Sichel D. E. (1993), "Business Cycle Asymmetry: a Deeper Look", *Economic Inquiry*, 31, 224-236.

Teräsvirta, T. (1998), "Modelling Economic Relationships with Smooth Transition Regressions", in *Handbook of Applied Economic Statistics*, eds. A. Ullah and D.E.A. Giles, New York: Marcel Dekker, 507-552.

Figure 1: US. Industrial Production Index, Textile, 1947.1-1996.4. Leave-k-out diagnostics. The first panel is a plot of the logarithms of the original series. The subsequent panels display the leave-k-out statistic $\hat{\tau}_{(I)}$, for k ranging from 1 to 5.